DOCUMENT RESUME

ED 472 164 TM 034 718

AUTHOR	McMillan, James H.
TITLE	The Relationship between Instructional and Classroom Assessment Practices of Elementary Teachers and Student Scores on High-Stakes Tests.
PUB DATE	2003-01-00
NOTE	18p.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; *Educational Practices; Elementary Education; *Elementary School Teachers; *High Stakes Tests; *Scores; Self Report; *Student Evaluation; Teaching Methods

ABSTRACT

This study sought to determine the relationships between teacher self-reported instructional and classroom assessment practices and scores on a state high-stakes test. Seventy-nine teachers participated. Average mathematics and reading test scale scores of students in each class were used as dependent variables, using a measure of aptitude as a covariate. Overall, there were few relationships, suggesting that many variations in instruction and assessment are related to high achievement. There was some evidence to suggest that use of cooperative learning and small groups, direct teaching, the use of formative assessment, and use of essay tests showed small positive relationships to achievement. Few differences were noted between mathematics and reading. Implications for improving external high-stakes tests are discussed. (Contains 3 tables and 12 references.) (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 472 164

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. McMillan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The Relationship Between Instructional and Classroom Assessment Practices of Elementary Teachers and Student Scores on High-Stakes Tests

James H. McMillan

Virginia Commonwealth University

Manuscript submitted for publication in *ERIC, Clearinghouse on Assessment and Evaluation*

January 8, 2003

BEST COPY AVAILABLE

TM034718

Abstract

The Relationship Between Instructional and Classroom Assessment Practices of Elementary Teachers and Student Scores on High-Stakes Tests

This study sought to determine the relationships between teacher self-reported instructional and classroom assessment practices and scores on a state high-stakes test. Seventy-nine fifth teachers participated. Average mathematics and reading test scale scores of students in each class were used as dependent variables, using a measure of aptitude as a covariate. Overall, there were few relationships, suggesting that many variations in instruction and assessment are related to high achievement. There was some evidence to suggest that use of cooperative learning and small groups, direct teaching, the use of formative assessment, and use of essay tests showed small positive relationships to achievement. Few differences were noted between mathematics and reading. Implications for improving external high-stakes tests are discussed.

As the importance of large-scale assessments has risen, so too has the impact of these tests on classroom teaching and assessment practices. While many believe that the consequences have been positive, previous research supports the conclusion that there are significant negative effects of high-stakes testing on teaching and learning. As noted by assessment expert Lorrie Shepard, "it is important to recognize the pervasive negative effects of accountability tests and the extent to which externally imposed testing programs prevent and drive out thoughtful classroom practices" (Shepard, 2000, p.9). Preliminary research on the consequences of high-stakes testing has suggested that it may de-professionalize teaching, increase rote memorization, narrow the curriculum, promote classroom assessment practices that mirror the format of the tests, and encourage a direct style of teaching (Darling-Hammond, 1988; McMillan, 2001; Shepard, 2000), but these findings are based mostly on national-level standardized tests. Cizek (2001), on the other hand, suggests several positive consequences, including increased use of student performance data to evaluate programs, increased knowledge about testing, renewed interest in teacher professionalism, and increased student learning.

It has been noted that the objective format and psychometric principles of large-scale testing conflict with implications for teaching and learning derived from contemporary views of learning represented by cognitive and constructivist paradigms (McMillan, in press; Shepard, 2000). As a result, teachers' decision-making in the classroom may be conflicted. What constructivist theories promote about student learning and teaching, such as authentic learning, deep understanding, intrinsic motivation, and constructed-response and student formative self-assessment, tend to conflict with pressures to align teaching and classroom assessments with external tests that emphasize simple understanding, decontextualized tasks, and selected-response tests. Some have suggested that high-stakes tests have influenced teachers' decisions to leave the profession.

While a number of studies document these impacts, most evidence is anecdotal, and rarely has there been in-depth research to fully understand how methods of instruction, classroom assessment, and student performance on high-stakes tests are related. Too often, what are voiced are purported positive or negative relationships, often depending on political or philosophical positions, without solid evidence. It is often argued that teachers are not mandated to use particular instructional strategies, that accountability is only for student outcomes, but there is little evidence about how teachers effectively incorporate instructional approaches with the pressures of external testing.

In an early study of the effects of high-stakes testing on instruction, Smith (1991) found that teachers used more worksheets and less hands-on instruction. A survey of 236 elementary teachers in North Carolina (Jones, Jones, Hardin, Chapman, Yarbrough, & Davis, 1999) found that 67% had changed their instructional methods as a result of high-stakes testing. The nature of the change was mixed, with approximately equal numbers using more and fewer inquiry projects and worksheets, with increases in using hands-on activities, group activities, and student-centered instruction. Firestone, Monfils, and Camilli (2001) conducted a survey of 287 fourth grade teachers in New Jersey and found that state-level testing and accountability demands were not sufficient to impact teaching and that most teachers have not dramatically changed practices. Rather, changes in instruction depended on local support and pressure.

What is lacking in most of the past research on the effects of high-stakes testing is empirical evidence that relates instructional and classroom assessment practices to actual test scores. The current study examines these relationships using test data for students in 79 Virginia elementary classrooms. It is also clear that teachers and administrators are pressured to raise test scores, but there is no evidence that particular instructional or classroom assessment strategies will increase accountability test scores. This study takes one step in investigating this issue.

Furthermore, it is important to account for student ability in examining standardized test score data so that established relationships are not confounded by entering student achievement and ability.

Two specific research questions were investigated:

What is the relationship between instructional methods emphasized and Standards of Learning (SOL) test scores?

What is the relationship between classroom assessment practices and SOL test scores?

Methodology

Sample

The convenience sample included 79 fifth grade teachers from 29 K-5 elementary schools in a suburban school district. The district is socially and economically diverse.

Instruments

Instruments for the study included the SOL measures for the dependent variables, Stanford Achievement Test scores as a measure of student ability, and a teacher survey of instructional and classroom assessment practices. In Virginia, SOL tests are administered to every fifth grade student in May. This study utilized the math and reading/language arts tests, which are separately administered 50 item multiple choice tests. The average scale score of fifth grade teachers for math and reading/language arts was calculated and used in the analyses. Stanford 9 reading and math scores for each student were obtained during the fall of the fourth grade and averaged for the fifth grade students in each class to provide a proxy for student ability.

The survey data were collected by teacher self-report in early June. There were six items that measured instructional practices and 13 items that focused on classroom assessment practices. The six instructional practice items were based in part on scales derived from research by Monfils, Camilli, Firestone, and Mayrowetz (2000), in which validity evidence was found to support

collaborative learning, active learning (constructivist approaches), and traditional teaching (direct instruction and independent seat work) as separate instructional components. Two additional items were added to the four generated from this research to focus on whether instruction more generally was focused on the state standards and SOL tests. Additional survey items on assessment were based on earlier research by McMillan (2002) that focused on the classroom assessment practices of elementary teachers. In that study evidence for validity supported four types of assessment practices (objective, essay, portfolios, and authentic), different cognitive levels assessed (recall or deep understanding), the extent to which teachers constructed their own assessments or used assessments supplied to them, and whether classroom assessments were aligned with the SOL tests.

The Likert-type scale used in McMillan (2002) was modified slightly for all questions. Teachers were asked to indicate the extent to which they had used each practice, separately for math and English/language arts.

Findings

The percentages of teachers responding to each point on the scale, means, and standard deviations are summarized in Tables 1 and 2. For English/language arts, direct instruction was clearly used most, with mean of 4.22 and 83% of the teachers indicating that they used this approach "quite a bit" or "extensively." Cooperative and small group activities and constructivist teaching methods were used more than independent seat or class work, with means of 3.96, 3.83and 3.67, respectively, with only slightly lower percentages of teachers using these approaches "quite a bit" or "extensively." As would be expected, instruction was focused heavily on the SOL and SOL tests. The results for math instruction were essentially the same.

For English/language arts assessment practices objective tests are used much more than essay, informal, performance, authentic, or portfolios (means 4.54, 4.01, 3.58, 3.33, 3.47, and 2.78, respectively). Teachers emphasized assessments that measure recall knowledge and deep understanding about the same, with more than 80% responding "quite a bit" or "extensively." Only

61% indicated "quite a bit" or "extensively" for assessments requiring student explanations. Classroom assessments were heavily focused on the SOL and SOL tests. Similar patterns of results were found in math, with the exception of essays, which were used much less.

Table 3 presents partial correlations between teacher responses and SOL test score results, with Stanford 9 scores used as covariates. For English/language arts only one correlation was statistically significant, which showed a positive relationship between the extent to which cooperative and small group activities were used and SOL scores (r=.34). The emphasis on direct instruction approached significance (r=.22). Other instructional variables were unrelated to SOL scores. In math, none of the instructional variables were related to SOL scores.

For classroom assessment practices none of the relationships were statistically significant. Two positive correlations approached significance for English/language arts, the extent to which essay tests and informal assessments were used. In math, extent of use of essay exams approached a positive statistically significant relationship, while the use of supplied assessments approached a significant negative correlation.

Two factors make it difficult to obtain significant relationships in this study. First, the high mean scores and small standard deviations on about half the items resulted in restricted ranges that make it difficult to establish statistically significant correlations. Given this limitation and the moderate sample size it is likely that there are some true relationships that were unable to be documented. Second, high correlations existed between the Stanford 9 and SOL scores (English .86; math .84). This suggests that a high percentage of the variability in SOL scores is accounted for by student ability, leaving little variation that can be explained by instructional and classroom assessment practices.

Conclusions

This study is limited to teacher perceptions of classroom practices and sample characteristics of primarily a suburban school district. It is possible that observational data would provide different results since some teachers may want to respond in socially desirable ways. Even though the survey was anonymous, the perceptions may not reflect realities of what has occurred in the classes.

Given this limitation, the findings suggest that while there may be some relationships between instructional and classroom assessment practices and student achievement on high-stakes tests, these relationships are both few and small. Many of the factors did not show any relationships with test scores, which suggests that variations in instructional practices and classroom assessments may not be responsible for differences on test scores. This finding is consistent with Firestone et al. (2001), and may mean that high high-stakes test results can be achieved with a variety of teaching methods and assessments. It may be that the quality of the delivery of different practices is more important than the approach is direct and assessment. That is, there may be several methods or approaches that can be used to result in high scores.

There were also some trends in the relationships that are important. Consistent with much research, there was a positive relationship between the use of cooperative and small group instruction and English test scores. This may mean that teachers who tend to use these techniques more will have higher test scores. The trend that showed a positive relationship between direct teaching and test scores supports what many see as a detrimental impact of objective high stakes tests. For classroom assessment practices, it is interesting that more use of essay tests for both English and math was related to higher objective test scores. It may be that essay tests require student learning that is consistent with the cognitive level of the tests. Even with the statistical adjustment for ability, it may also be that students in classes that emphasize essays more are in general more capable than students in classes that use less essay assessment.

The finding of a positive relationship between the use of informal, formative assessment and test scores is consistent with recent research reported by Black & Wiliam (1998). As pointed out by Stiggins (2002) and Brookhart (2001), teacher use of formative assessments and frequent, specific and descriptive feedback to students, is supported by recent cognitive learning and motivation theories. Even though these findings are primarily correlational, this may suggest an important way for teachers to have direct control on a factor that may enhance high-stakes test results.

What can teachers and administrators do to improve students' scores on high-stakes tests? This study suggests that moderate impacts can be made by some practices, but much research is needed to establish relationships between instructional and assessment practices and test scores. Specifically, there is a need to measure both instructional and assessment practices in ways that provide more variability so that there is greater sensitivity to measure relationships. There is also a need to provide better measures of student ability so that the unique contributions of teaching and assessment can be determined.

References

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education, 8*(2), 153-169.

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20*(4), 19-27.

Darling-Hammond, L. (1988). Accountability and teacher professionalism. *American Educator, 12*, 8-13.

Firestone, W. A., Monfils, L., & Camilli, G. (2001). Pressure, support, and instructional chante in the context of a state testing program. Paper presented at the annual meeting of the American Educational Research Association, Seattle.

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yargrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 80*, 199-203.

McMillan, J. H. (2001). High stakes testing. Issue brief. Richmond, VA: Commonwealth Educational Policy Center (www.edpolicyvcu.org).

McMillan, J. H. (2002). Elementary teachers' classroom assessment and grading practices. *Journal of Educational Research, 95*(4), 203-213.

McMillan, J. H. (in press). Understanding and improving teachers' classroom assessment decision-making: Implications for theory and practice. *Educational Measurement: Issues and Practice.*

Monfils, L., Camilli, G., Firestone, W., & Mayrowetz, D. (2000). Multidimensional analysis of scales developed to measure standards-based instruction in response to systemic reform. Rutgers, The State University of New Jersey.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*

29(7), 4-14.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning, *Assessment in Education, 5*(1), 103-110.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan, 83*(10). 758-765.

Table 1

Percentages (rounded), Means and Standard Deviations of Fifth Grade Teacher Responses To English/Language Questions

N=79

To what extent did you use:	Not at All	Very Little	Some	Quite a Bit	Extensively	Mean	S.D.
1. assessments that measure student recall knowledge or simple comprehension	---	3	16	29	53	4.32	.83
2. assessments that measure student deep understanding, reasoning, and/or application	---	---	16	43	42	4.26	.72
3. objective assessments (e.g., multiple choice, matching, short answer)	---	1	6	29	63	4.54	.68
4. assessments supplied to you (e.g., from publishers, software, instructor's guide, division)	3	17	22	27	31	3.68	1.16
5. assessments constructed by yourself	---	1	13	41	45	4.29	.74
6. performance assessments (e.g., structured teacher observations or ratings of performance such as a project, speech or paper)	1	9	39	33	18	3.58	.93
7. essay-type assessments	4	11	46	27	13	3.33	.97
8. portfolios	15	27	34	13	11	2.78	1.20
9. direct instruction (i.e., structured, systematic teaching in	---	4	13	41	42	4.22	.82

which information is presented to students with review and practice; teacher-centered)							
10. assessments aligned closely to the SOL goals and objectives	---	1	---	11	87	4.85	.46
11. assessments aligned closely to SOL test(s)	---	1	8	22	69	4.59	.69
12. informal assessments that provide immediate feedback to students	---	---	27	44	29	4.01	.75
13. constructivist teaching methods (e.g., active learning, contextualized, creating personal meaning for students, student-centered, discovery, reflection)	1	7	29	35	29	3.83	.97
14. instruction focused on the SOL goals and objectives	---	---	---	17	84	4.84	.37
15. instruction focused on the SOL test(s)	1	---	11	20	67	4.52	.80
16. assessments or instruction that required student explanations of work, conclusions, opinions, ideas, and/or answers	4	5	30	41	20	3.68	.98
17. authentic assessments (i.e., questions based on "real world" problems or materials)	3	10	32	49	6	3.47	.86
18. cooperative and/or small group learning activities	---	3	28	41	29	3.96	.82
19. independent seat or class work with assignments and/or worksheets	---	11	33	33	23	3.67	.96

Table 2

Percentages (rounded), Means and Standard Deviations of Fifth Grade Teacher Responses To Math Questions

N=79

To what extent did you use:	Not at All	Very Little	Some	Quite a Bit	Exten-sively	Mean	S.D.
1. assessments that measure student recall knowledge or simple comprehension	---	3	15	29	54	4.33	.83
2. assessments that measure student deep understanding, reasoning, and/or application	---	---	10	41	49	4.39	.67
3. objective assessments (e.g., multiple choice, matching, short answer)	1	1	4	30	63	4.52	.77
4. assessments supplied to you (e.g., from publishers, software, instructor's guide, division)	3	9	23	26	39	3.90	1.11
5. assessments constructed by yourself	---	3	16	39	43	4.21	.82
6. performance assessments (e.g., structured teacher observations or ratings of performance such as a project, speech or paper)	3	14	40	24	19	3.41	1.04
7. essay-type assessments	16	27	34	14	9	2.73	1.15
8. portfolios	27	29	24	14	6	2.43	1.20
9. direct instruction (i.e., structured, systematic teaching in which information is presented to students with	---	4	14	39	43	4.20	.84

review and practice; teacher-centered)							
10. assessments aligned closely to the SOL goals and objectives	---	---	---	9	92	4.92	.28
11. assessments aligned closely to SOL test(s)	---	---	10	20	70	4.61	.67
12. informal assessments that provide immediate feedback to students	---	---	20	44	36	4.16	.74
13. constructivist teaching methods (e.g., active learning, contextualized, creating personal meaning for students, student-centered, discovery, reflection)	---	7	30	35	28	3.83	.92
14. instruction focused on the SOL goals and objectives	---	---	---	14	86	4.86	.35
15. instruction focused on the SOL test(s)	---	---	11	20	69	4.58	.69
16. assessments or instruction that required student explanations of work, conclusions, opinions, ideas, and/or answers	1	7	35	38	18	3.65	.91
17. authentic assessments (i.e., questions based on "real world" problems or materials)	1	6	27	54	13	3.70	.82
18. cooperative and/or small group learning activities	---	3	28	38	31	3.97	.85
19. independent seat or class work with assignments and/or worksheets	---	13	31	34	23	3.66	.97

Table 3

Partial Correlation Coefficients of the Relationship Between Fifth Grade Classroom Assessment and Instructional Practices and SOL Test Scores, Controlling for Student Ability

N=79

Assessment or Instructional Practice	SOL Test	
	English	Math
1. assessments that measure student recall knowledge or simple comprehension	.10	.03
2. assessments that measure student deep understanding, reasoning, and/or application	.10	.13
3. objective assessments (e.g., multiple choice, matching, short answer)	.15	.02
4. assessments supplied to you (e.g., from publishers, software, instructor's guide, division)	-.04	-.26*
5. assessments constructed by yourself	-.02	.18
6. performance assessments (e.g., structured teacher observations or ratings of performance such as a project, speech or paper)	-.03	-.11
7. essay-type assessments	.25*	.24*
8. portfolios	-.01	.03

9. direct instruction (i.e., structured, systematic teaching in which information is presented to students with review and practice; teacher-centered)	.22*	-.11
10. assessments aligned closely to the SOL goals and objectives	.22*	-.12
11. assessments aligned closely to SOL test(s)	-.06	.00
12. informal assessments that provide immediate feedback to students	.23*	.02
13. constructivist teaching methods (e.g., active learning, contextualized, creating personal meaning for students, student-centered, discovery, reflection)	.03	.00
14. instruction focused on the SOL goals and objectives	.16	.00
15. instruction focused on the SOL test(s)	.03	-.02
16. assessments or instruction that required student explanations of work, conclusions, opinions, ideas, and/or answers	.19	.15
17. authentic assessments (i.e., questions based on "real world" problems or materials)	.12	.07
18. cooperative and/or small group learning activities	.34**	-.01
19. independent seat or class work with assignments and/or worksheets	-.05	-.16

* $p \leq .10$; ** $p \leq .01$

TM034718

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: The Relationship Between Instructional and Classroom Assessment Practices of Elementary Teachers and Student Scores on High-Stakes Tests

Author(s): James H. McMillan

Corporate Source: Virginia Commonwealth University

Publication Date: Jan 8, 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	**Level 2A**	**Level 2B**
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: James McMillan	Printed Name/Position/Title: James McMillan / Professor	
Organization/Address: Virginia Commonwealth U. Box 842020 Richmond, VA 23284-2020	Telephone: 804 828-1332 x553	Fax: 804 225-3554
	E-mail Address: jmcmillan@vcu.edu	Date: 1/8/03

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: info@ericfac.piccard.csc.com
WWW: http://ericfacility.org